1. **Descriptive Statistics – Bivariate & Multivariate: [60 marks]**

**Provide all the code (.R file) and all results from R along with your interpretation in a document.**

The Melanoma data frame has data on 205 patients in Denmark with malignant melanoma. The data is stored in R in the package MASS, in a dataset called ‘Melanoma’.

1. Create bivariate plots/tables to explore if there is a relationship between 'status' and all other variables. Interpret each plot/table.

status\_time\_plot = plot(status,time, xlab="Status",

main="Status VS Time",col=3:5)

legend("topleft",legend=c("Dead","Alive", "Dead\_Other"),fill=3:5)

Chart, box and whisker chart

Description automatically generated

There are more patient alive than dead from melanoma and other dead has Alive media is higher than the other, the relationship between dead and other\_dead is that there are patient that died from other cause than from melanoma.

status\_age\_plot = plot(status,age, xlab="Status",

main="Status VS Age",col=3:5)

legend("topleft",legend=c("Dead","Alive", "Dead\_Other"),fill=3:5)

Chart, box and whisker chart

Description automatically generated

The dead patient from both death from melanoma and other are relatively older than the alive patient as the centre line fall bellows both dead conditions. The fact that dead from other centre line is higher shows that it can be said that age has more effect status than the melanoma.

status\_year\_plot = plot(status,year, xlab="Status",

main="Status VS Year",col=3:5)

legend("topleft",legend=c("Dead","Alive", "Dead\_Other"),fill=3:5)

Chart, box and whisker chart

Description automatically generated

As the status plot overlap shows that the year the operation was done has no effect on status.

status\_year\_plot = plot(status,thickness, xlab="Status",

main="Status VS Thickness",col=3:5)

legend("topleft",legend=c("Dead","Alive", "Dead\_Other"),fill=3:5)

Chart, box and whisker chart

Description automatically generated

The plot shows that there is slight corelation between the thickness of the tumour and status for dead my melanoma, has the patent the larger tumour are on the dead status group.

status\_sex\_add=addmargins(table(status,sex))

Table

Description automatically generated with low confidence

Sex has no corelation to the status.

status\_ulcer\_add =addmargins(table(status,ulcer))

Table

Description automatically generated with low confidence

Ulcer has a corelation in the patent status.

1. Create suitable scatterplots to analyse all relationships between the continuous numeric variables. Interpret each scatterplot
2. Using the scatterplots from part b, investigate if there is a different relationship

between the numeric variables depending on 'status'. Interpret each scatterplot.

plot(time,age,main="Survival time in days(Time) VS Patients Age in Year (Age)")

Chart, scatter chart

Description automatically generated

The older the patent are the less survival time they have; this is normal and not an effect of the melanoma.

plot(time,thickness,main="Survival time in days(Time) VS Tumour Thickness in mm (Thickness)")

Chart, scatter chart

Description automatically generated

There is a trend in the size of the tumour to the length of time of survival. The tumour thickness has a corelation to survival time, this is also reflected in the Status vs tumour plot.

plot(age,thickness,main="Patients Age in Year (Age) VS Tumour Thickness in mm (Thickness)")

Chart, scatter chart

Description automatically generated

There is a trend in the age vs tumour size as the larger tumour size are in the older patients.

plot(age,time,main="Patients Age in Year (Age) VS Survival time in days(Time)")

Chart, scatter chart

Description automatically generated

Age always have effect on survival time this is also shown in the status VS age table.

1. Change the status, ulcer, and sex variables to factor variables with meaningful labels and create a flatted 3-way table to explore the relationship between all three variables together. Interpret this table.

A picture containing text

Description automatically generated

In the dead vs ulcer, there is more dead status patent with ulcer precent than in absent. Sex has no effect on the distribution.

1. **Data Camp: [40 marks]**

**Provide screenshots/pictures of the XP points completed for each chapter.**

* Complete the next three chapters in Introduction to R course in Data Camp:
  + 1. Factors

Graphical user interface, text, application, email

Description automatically generated

* + 1. Data frames

Table

Description automatically generated

* + 1. Lists

Graphical user interface, text

Description automatically generated with medium confidence